

Implementasi *Association Rules* dengan Algoritma Apriori pada Dataset Kemiskinan

Winda Aprianti¹, Khairul Anwar Hafizd², M. Redhy Rizani³

^{1,2,3}Jurusan Teknik Informatika, Politeknik Negeri Tanah Laut

¹winda@politala.ac.id

Abstrak

Usaha pengentasan kemiskinan terus dilakukan di Kabupaten Tanah Laut. Untuk membantu pemerintah dalam perumusan kebijakan pengentasan kemiskinan maka diperlukan pengetahuan mengenai indikator yang berkaitan dengan kemiskinan dan bagaimana indikator-indikator tersebut saling mempengaruhi. Tujuan penelitian ini adalah menerapkan *Association rules* dengan algoritma Apriori pada dataset kemiskinan untuk mencari pola hubungan antar indikator. Dataset yang terdiri dari 46 atribut merupakan data sekunder BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan tahun 2010-2014. Hasil penerapan *association rules* dengan algoritma apriori menggunakan *minimum support* 30% dan *minimum confidence* 80% menghasilkan 4614 *rules* hubungan antar indikator.

Katakunci: *Association Rules, Algoritma Apriori, Indikator Kemiskinan*

1 Pendahuluan

Kemiskinan masih menjadi permasalahan yang memerlukan perhatian lebih dari pemerintah. Kalimantan Selatan adalah salah satu Provinsi di Indonesia yang mengalami permasalahan dalam penanggulangan kemiskinan. Penurunan persentase kemiskinan di Kalimantan Selatan dari

tahun 2000 sampai dengan 2013 adalah sebesar 8.29%. Persentase kemiskinan di Kabupaten Tanah Laut sendiri telah menurun 5.71% dari tahun 2000-2013 [1].

Sekarang ini, kemiskinan tidak lagi dipandang sebagai permasalahan rendahnya pendapatan seseorang, melainkan disebabkan oleh beragamnya faktor penyebab kemiskinan [2, 3, 4]. Faktor pengangguran dan Produk Domestik Regional Bruto (PDRB) termasuk yang diteliti terhadap kemiskinan pada penelitian Wijayanto (2010), Rusdati dan Sebayang (2013), tetapi memberikan hasil yang berbeda.

Hal ini menunjukkan bahwa permasalahan kemiskinan sangat kompleks dan perbedaan karakteristik wilayah, serta teknik yang digunakan juga berpengaruh untuk mengetahui indikator penyebab kemiskinan dan perumusan kebijakan untuk pengentasan kemiskinan. Penelitian tentang kemiskinan sebelumnya menggunakan data numerik, tetapi dalam penelitian ini data kemiskinan ditransformasi menjadi data kategorikal. *Task* data mining yang digunakan untuk mencari pola hubungan antar atribut adalah *Association Rules* dengan Algoritma Apriori. Penggunaan Algoritma Apriori dinilai efektif untuk menemukan pola hubungan antara beberapa atribut dengan tingkat kecelakaan lalu lintas, hubungan lingkungan fisik dengan terjadinya kebakaran hutan, serta pola pergerakan harga saham antar sektor dan antar perusahaan [5, 6, 7]. Berdasarkan penelitian terdahulu, maka pada penelitian ini *Association Rules* dengan Algoritma Apriori diimplementasikan pada dataset kemiskinan.

2 Tinjauan Pustaka

2.1 Penelitian Terkait

Sitanggang (2013) menerapkan *association rules* dengan Algoritma Apriori pada dataset kebakaran hutan yang berisi data pada lingkungan fisik (tutupan lahan, sungai, jalan dan pusat kota), sosial-ekonomi (sumber pendapatan, populasi, dan jumlah sekolah), cuaca (hujan, angin kecepatan, dan temperatur layar), dan lahan gambut. Hasil penelitian mengungkapkan

324 aturan asosiasi multidimensi yang menunjukkan hubungan antara hotspot kejadian dan faktor lainnya. Asosiasi antara hotspot terjadinya dengan objek geografis lainnya ditemukan untuk *minimum support* 10% dan *minimum confidence* 80% [6]. Algoritma apriori juga digunakan untuk menentukan pola hubungan antara usia, jenis kecelakaan, waktu terjadinya kecelakaan, kepemilikan Surat Ijin Mengemudi (SIM), jenis kelamin, dan pekerjaan dengan tingkat kecelakaan kecelakaan lalu lintas di Jalan Raya Kabupaten Sleman [5]. Penggunaan variabel *fuzzy* sebelum penerapan Algoritma Apriori diterapkan pada dataset harga saham untuk mengetahui pola pergerakan harga saham antar perusahaan [7].

2.2 Association Rules

Association rules adalah salah satu *task* data mining deskriptif yang bertujuan untuk menemukan aturan asosiasi antara item-item data. Langkah utama yang perlu dalam *association rules* adalah mengetahui seberapa sering kombinasi item muncul dalam *database*, yang disebut sebagai *frequent patterns* [8]. Pramudiono dalam [7] menyatakan bahwa penting tidaknya suatu aturan asosiatif dapat diketahui dengan dua parameter, *support* yaitu persentase kombinasi item dalam *database* dan *confidence* yaitu kuatnya hubungan antar item dalam aturan asosiatif.

Support dan *confidence* dituliskan sebagai Persamaan 1 dan Persamaan 2 [8].

$$\text{support}(A \Rightarrow B) = P(A \cup B) \quad (1)$$

$$\text{confidence}(A \Rightarrow B) = P(B | A) \quad (2)$$

Jika *support itemset* dari *itemset I* memenuhi *minimum support threshold* yang sudah ditentukan maka *I* adalah *frequent k-itemset*. Secara umum *frequent k-itemset* dilambangkan dengan L_k . Berdasarkan Persamaan (2) diperoleh

$$\text{confidence}(A \Rightarrow B) = P(B|A) = \frac{\text{support}(A \cup B)}{\text{support}(A)} \quad (3)$$

2.3 Algoritma Apriori

Algoritma Apriori adalah salah satu algoritma untuk melakukan pencarian *frequent itemset* dengan *association rules*. Algoritma Apriori menggunakan pendekatan *level-wise search*, dimana *k-itemset* digunakan untuk memperoleh $(k+1)$ -*itemset*. Proses ini dilakukan hingga tidak ada lagi kombinasi yang dapat dibentuk[8]. Berdasarkan persamaan (1), (2), dan (3), serta pembahasan pada [9], diperoleh Persamaan (4), (5), dan (6).

Persamaan (4) digunakan untuk menghitung nilai *support* untuk sebuah *item*.

$$\text{Support } (A) = \frac{\text{Jumlah Transaksi yang mengandung } A}{\text{Total Transaksi}} \quad (4)$$

Persamaan (5) digunakan untuk menghitung nilai *support* dari 2 *item*.

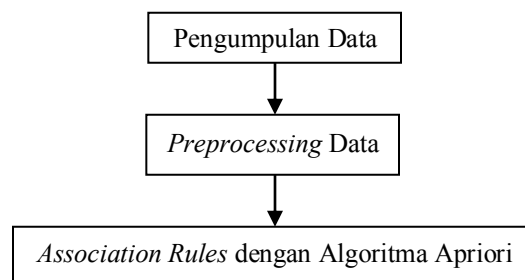
$$\text{Support } (A, B) = \frac{\text{Jumlah Transaksi yang mengandung } A \text{ dan } B}{\text{Total Transaksi}} \quad (5)$$

Persamaan (6) digunakan untuk menghitung nilai *confidence* dari *rule* $A \rightarrow B$.

$$\text{Confidence } (A \rightarrow B) = \frac{\text{Jumlah Transaksi yang mengandung } A \text{ dan } B}{\text{Jumlah Transaksi yang mengandung } A} \quad (6)$$

3 Metodologi Penelitian

Metodologi penelitian disajikan pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

3. 1 Pengumpulan Data

Dataset yang digunakan merupakan data sekunder dari BPS Kabupaten Tanah Laut dan BPS Provinsi Kalimantan Selatan Tahun 2009-2014. Adapun dataset yang berjumlah 46 atribut ini terdiri dari:

1. Jumlah penduduk miskin (A1).
2. Pengeluaran rata-rata perkapita untuk kelompok makanan (A2).
3. Pengeluaran rata-rata perkapita untuk kelompok non makanan (A3).
4. Produk Domestik Regional Bruto (PDRB) yang terbagi menjadi Atas Dasar Harga Berlaku (ADHB) dan Atas Dasar Harga Konstan (ADHK) pada setiap sektornya sejumlah 14 sektor, yaitu PDRB Sektor Bangunan; PDRB Sektor Industri Pengolahan; PDRB Sektor Jasa-Jasa; PDRB Sektor Keuangan; PDRB Sektor Persewaan dan Jasa Perusahaan; PDRB Sektor Listrik dan Gas; PDRB Sektor Air Bersih; PDRB Sektor Pengangkutan; PDRB Sektor Komunikasi; PDRB Sektor Perdagangan, Hotel dan Restoran; PDRB Sektor Pertambangan dan Penggalan; PDRB Sektor Pertanian dan Tanaman Pangan; PDRB Sektor Perkebunan; PDRB Sektor Peternakan; PDRB Sektor Kehutanan; serta PDRB Sektor Perikanan (A4 sampai dengan A35).
5. Inflasi (A36).
6. Luas panen tanaman padi dan palawija yang terdiri dari 7 kelompok, yaitu padi; jagung; kedelai; kacang tanah; kacang hijau; ubi kayu; dan ubi jalar (A37 sampai dengan A43).
7. Jumlah industri manufaktur besar dan sedang (A44).
8. Jumlah pengangguran (A45).
9. Jumlah penduduk (A46).

Semua atribut pada dataset merupakan data numerik. Namun data inflasi dan luas panen tanaman berupa data bulanan, sedangkan data yang lain berupa data tahunan. Sehingga data inflasi dan luas panen tanaman diubah menjadi data tahunan dengan cara menghitung rata-rata dari masing-masing data tersebut.

3. 2 Preprocessing Data

Tahapan *preprocessing* data dibagi menjadi 2 tahap, yaitu *cleaning* data dan transformasi data kategorikal.

Cleaning data untuk menghilangkan data yang tidak konsisten. Atribut ADHB (A6) dan ADHK (A22) Sektor Jasa-jasa untuk Tahun 2009-2010 dan Tahun 2011-2014 mempunyai pengumpulan data jasa-jasa yang berbeda sehingga kedua atribut harus dihilangkan dari dataset. Selain kedua atribut tersebut, atribut jumlah industri besar dan sedang (A44) juga perlu dihilangkan dari dataset karena data untuk tahun 2011-2014 tidak diketahui. Sehingga dataset sekarang berjumlah 43 atribut.

Berdasarkan kajian pada [7], maka proses transformasi data dimulai dengan melakukan perhitungan selisih dari data yang telah terurut dan mencari persentase data. Hasil persentase data kemudian dibagi menjadi 3 bagian yaitu kenaikan, tetap, dan penurunan. Untuk pergerakan yang tidak berubah diberikan kategori Tetap. Sedangkan untuk menentukan kategori kenaikan dan penurunan, data difuzzifikasi menggunakan fungsi keanggotaan segitiga tipe bahu untuk nilai persentase kenaikan dan penurunan. Setiap fungsi keanggotaan kenaikan dan penurunan dibagi menjadi 3 kategori, yaitu Rendah, Sedang, dan Tinggi. Hal ini menyebabkan setiap atribut numerik mempunyai kemungkinan berubah menjadi 7 data kategorikal, sehingga dataset yang awalnya berukuran 43×5 sekarang menjadi dataset berukuran 301×5 .

3. 3 Association Rules dengan Algoritma Apriori

Setelah dataset ditransformasi menjadi atribut kategorikal, berikutnya diterapkan *association rules* dengan algoritma Apriori. Langkah-langkah pada tahap ini adalah sebagai berikut.

1. Menentukan *minimum support* dan *minimum confidence*.
2. Menghitung *k-itemset*, dengan *k* adalah jumlah kombinasi *itemset*.

3. Menentukan *frequent(k) itemset*, yaitu *itemset* yang memenuhi *minimum support*.
4. Menentukan *rules*.

4 Hasil Penelitian

Dataset hasil *preprocessing* yang menghasilkan 43 atribut numerik disajikan pada Tabel 1.

Tabel 1. Dataset Hasil *Preprocessing* Data

No	Nama Atribut	Atribut	Tahun					
			2009	2010	2011	2012	2013	2014
1	Penduduk Miskin	A1	14039	13654	14104	14698	15218	13856
2	Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan	A2	308434	345529	364140	424891	474028	474164
3	Pengeluaran Rata-Rata per Kapita untuk Kelompok Non Makanan	A3	264873	242825	326409	383895	370033	369209
...
41	Luas Panen Tanaman Ubi Jalar (Ha)	A43	151	174	180	203	205	228
42	Jumlah Pengangguran	A45	7992	5991	7962	6519	4589	4775
43	Jumlah Penduduk	A46	274526	297814	303190	308510	131725	319098

Dataset pada Tabel 1 kemudian ditransformasi menjadi data kategorikal. Apabila hasil selisih data menunjukkan kenaikan maka digunakan fungsi keanggotaan *fuzzy* yang disajikan pada Persamaan 4, 5, dan 6.

Fungsi keanggotaan untuk kategori Naik Rendah (NR):

$$\mu_{NR}(x) = \begin{cases} 1 & , x \leq 33.33 \\ \frac{40-x}{6.67} & , 33.33 < x < 40 \\ 0 & , x \geq 40 \end{cases} \quad (4)$$

Fungsi keanggotaan untuk kategori Naik Sedang (NS):

$$\mu_{NS}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-33.33}{16.67} & , 33.33 \leq x < 50 \\ \frac{66.67-x}{16.67} & , 50 < x < 66.67 \\ 0 & , x \geq 66.67 \end{cases} \quad (5)$$

Fungsi keanggotaan untuk kategori Naik Tinggi (NT):

$$\mu_{NT}(x) = \begin{cases} 0 & , x \leq 33.33 \\ \frac{x-60}{6.67} & , 60 < x < 66.67 \\ 1 & , x \geq 66.67 \end{cases} \quad (6)$$

Sedangkan apabila hasil selisih data menunjukkan penurunan maka digunakan Persamaan 7, 8, dan 9.

Fungsi keanggotaan untuk kategori Turun Rendah (TR):

$$\mu_{TR}(x) = \begin{cases} 1 & , x \geq -33.33 \\ \frac{x+40}{6.67} & , -40 < x < -33.33 \\ 0 & , x \leq -40 \end{cases} \quad (7)$$

Fungsi keanggotaan untuk kategori Turun Sedang (TS):

$$\mu_{TS}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{-x-33.33}{16.67} & , -50 \leq x < -33.33 \\ \frac{x+66.67}{16.67} & , -66.67 < x < -50 \\ 0 & , x \leq -66.67 \end{cases} \quad (8)$$

Fungsi keanggotaan untuk kategori Turun Tinggi (TT):

$$\mu_{TT}(x) = \begin{cases} 0 & , x \geq -33.33 \\ \frac{x+60}{6.67} & , -66.67 < x < -60 \\ 1 & , x \leq -66.67 \end{cases} \quad (9)$$

Hasil transformasi data disajikan pada Tabel 2.

Tabel 2. Dataset Hasil Transformasi Data

No	Atribut	Tahun					
		2009	2010	2011	2012	2013	2014
1	A1	TR_A1	NR_A1	NR_A1	NR_A1	TR_A1	TR_A1
2	A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2	NR_A2
3	A3	TR_A3	NR_A3	NR_A3	TR_A3	TR_A3	TR_A3
...
41	A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43	NR_A43
42	A45	TR_A45	NR_A45	TR_A45	TR_A45	NR_A45	TR_A45
43	A46	NR_A46	NR_A46	NR_A46	TS_A46	NT_A46	NR_A46

Pada Tabel 2, dataset telah berupa atribut kategori: Turun Rendah (TR), Turun Sedang (TS), Turun Tinggi (TT), Tetap (T), Naik Rendah (NR), Naik Sedang (NS), dan Naik Tinggi (NT). Dataset ini kemudian dibentuk menjadi *itemset* yang disajikan pada Tabel 3.

Tabel 3. *Itemset*

No	Item	waktu (t)					Jumlah
		t1	t2	t3	t4	t5	
1	NR_A1	0	1	1	1	0	3
2	NS_A1	0	0	0	0	0	0
3	NT_A1	0	0	0	0	0	0
...
299	TR_A46	0	0	0	0	0	0
300	TS_A46	0	0	0	1	0	1
301	TT_A46	0	0	0	0	0	0

Dengan *minimum support* = 30% dan *minimum confidence* = 80%, penerapan *association rules* dengan algoritma Apriori menghasilkan 1-*itemsets*, 2-*itemsets*, 3-*itemsets*, 4-*itemsets*, 5-*itemsets*, 6-*itemsets*, 7-*itemsets*, dan 8-*itemsets* seperti yang disajikan pada Gambar 2.

```

Associator output

Apriori
=====

Minimum support: 0.95 (5 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8
Size of set of large itemsets L(2): 28
Size of set of large itemsets L(3): 56
Size of set of large itemsets L(4): 70
Size of set of large itemsets L(5): 56
Size of set of large itemsets L(6): 28
Size of set of large itemsets L(7): 8
Size of set of large itemsets L(8): 1

Best rules found:

1. A4=NR_A4 5 ==> A2=NR_A2 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. A2=NR_A2 5 ==> A4=NR_A4 5 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)

```

Gambar 2. Hasil Penerapan *Association Rules* dengan Algoritma Apriori Menggunakan Weka

Setelah mendapatkan *itemsets*, maka selanjutnya diperoleh 4614 *rules* yang menunjukkan hubungan antar indikator (Tabel 4).

Tabel 4. *Rules* yang diperoleh

No	Rule
1	A4=NR_A4 ==> A2=NR_A2
2	A2=NR_A2 ==> A4=NR_A4
3	A5=NR_A5 ==> A2=NR_A2
...	...
4612	A5=NR_A5 A23=NR_A23 A25=NR_A25 A33=NR_A33 A43=NR_A43 ==> A29=NR_A29
4613	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A43=NR_A43 ==> A33=NR_A33
4614	A5=NR_A5 A23=NR_A23 A25=NR_A25 A29=NR_A29 A33=NR_A33 ==> A43=NR_A43

Pada Tabel 4 untuk *rule* 1 diperoleh pola hubungan antara indikator 4, yaitu ADHB Sektor Bangunan dengan indikator 2, yaitu Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan yang menunjukkan bahwa apabila ADHB Sektor Bangunan naik rendah maka Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan juga akan mengalami naik rendah. Pola hubungan yang diperoleh dari *rule* 1 ini dapat membantu pemerintah dalam pengambilan kebijakan yang berkaitan dengan ADHB Sektor Bangunan dan Pengeluaran Rata-Rata per Kapita untuk Kelompok Makanan dengan memperhatikan pola hubungan kedua atribut ini. Begitu juga dengan *rule* 2 sampai dengan *rule* 4614.

5 Kesimpulan

Berdasarkan hasil penerapan *association rules* dengan algoritma Apriori pada dataset kemiskinan menggunakan *minimum support* 30% dan *minimum confidence* 80% diperoleh 4614 *rules* hubungan antar indikator. Namun, banyaknya jumlah *rules* yang dihasilkan mengakibatkan banyaknya pengetahuan mengenai pola hubungan antar indikator. Pengetahuan yang terlalu luas ini berakibat pada kesulitan pengambilan keputusan untuk menentukan pola hubungan antar indikator mana yang lebih berpengaruh terhadap kemiskinan.

Guna menghasilkan pola hubungan indikator yang lebih spesifik maka akan diterapkan *Principal Component Analysis* sebagai salah satu *preprocessing* data pada penerapan Algoritma Apriori untuk penelitian selanjutnya.

6 Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Kemenristekdikti atas bantuan biaya pada skema Penelitian Dosen Pemula Tahun Anggaran 2017.

7 Pustaka

- [1] Badan Pusat Statistik (BPS), diakses dari <http://www.bps.go.id> pada tanggal 16 Maret 2016 jam 10.00 WITA.
- [2] Nurwati, N., 2008, “Kemiskinan: Model Pengukuran, Permasalahan, dan Alternatif Kebijakan”, *Jurnal Kependudukan Padjadjaran*, Vol. 10, No. 1, Januari 2008: 1-11.
- [3] Rusdarti dan Sebayang, K.L., 2013, “Faktor-faktor yang mempengaruhi tingkat kemiskinan di Provinsi Jawa Tengah”, *Jurnal Economia*, Volume 9, Nomor 1, April 2013.
- [4] Wijayanto, R. D., 2010, Analisis Pengaruh PDRB, Pendidikan, dan Pengangguran terhadap Kemiskinan di Kabupaten/Kota Jawa Tengah Tahun 2005 – 2008, Skripsi, Fakultas Ekonomi Universitas Diponegoro, Semarang.
- [5] Hakim, L. dan Fauzy, A., 2015, “Penentuan Pola Hubungan Kecelakaan Lalu Lintas Menggunakan Metode *Association Rules* dengan Algoritma *Apriori*”, *University Research Colloquium 2015*, ISSN 2407-9189.
- [6] Sitanggang, S. I., 2013, “Spatial Multidimensional Association Rules Mining in Forest Fire Data”, *Journal of Data Analysis and Information Processing*, 2013, 1, 90-96.
- [7] Arafah, A.A. dan Mukhlash, I., 2015, “The Application of Fuzzy Association Rule on Co-Movement Analyze of Indonesian Stock Price”, *Procedia Computer Science* 59 (2015) 235-243.
- [8] Han, J., Kamber, M., dan Pei, J., 2011, *Data Mining: Concepts and Techniques Third Edition*, Morgan Kaufmann, USA.
- [9] Kusriani dan Lutfi, E.T., 2009, *Algoritma Data Mining*, Andi, Yogyakarta.